

Python文件操作

张果

厦门大学 王亚南经济研究院

2016年9月25日

目录

1 os模块

2 os.path模块

3 一些实例

相关模块

- os: 文件系统的访问
- os.path: 管理和操作路径名

相关模块

- os: 文件系统的访问
- os.path: 管理和操作路径名

文件处理

- `os.remove(fname)` # 删除文件
- `os.rename(oldname,newname)` # 重命名/移动文件
- `os.walk(path)` # 生成一个目录树下所有文件

文件处理

- `os.remove(fname)` # 删除文件
- `os.rename(oldname,newname)` # 重命名/移动文件
- `os.walk(path)` # 生成一个目录树下所有文件

文件处理

- `os.remove(fname)` # 删除文件
- `os.rename(oldname,newname)` # 重命名/移动文件
- `os.walk(path)` # 生成一个目录树下所有文件

目录、文件夹

- `os.chdir(path)` # 改变当前工作目录
- `os.listdir(path)` # 列出指定目录的文件名
- `os.getcwd()` # 返回当前工作目录
- `os.mkdir(path)` # 创建目录

目录、文件夹

- `os.chdir(path)` # 改变当前工作目录
- `os.listdir(path)` # 列出指定目录的文件名
- `os.getcwd()` # 返回当前工作目录
- `os.mkdir(path)` # 创建目录

目录、文件夹

- `os.chdir(path)` # 改变当前工作目录
- `os.listdir(path)` # 列出指定目录的文件名
- `os.getcwd()` # 返回当前工作目录
- `os.mkdir(path)` # 创建目录

目录、文件夹

- `os.chdir(path)` # 改变当前工作目录
- `os.listdir(path)` # 列出指定目录的文件名
- `os.getcwd()` # 返回当前工作目录
- `os.mkdir(path)` # 创建目录

分隔

- `os.path.basename(fname)` # 去掉目录路径，返回文件名
- `os.path.dirname(fname)` # 去掉文件名，返回目录路径
- `os.path.join(dirname,basename)` # 组合目录名和文件名
- `os.path.split(fname)` # 拆分目录名和文件名

分隔

- `os.path.basename(fname)` # 去掉目录路径，返回文件名
- `os.path.dirname(fname)` # 去掉文件名，返回目录路径
- `os.path.join(dirname,basename)` # 组合目录名和文件名
- `os.path.split(fname)` # 拆分目录名和文件名

分隔

- `os.path.basename(fname)` # 去掉目录路径，返回文件名
- `os.path.dirname(fname)` # 去掉文件名，返回目录路径
- `os.path.join(dirname,basename)` # 组合目录名和文件名
- `os.path.split(fname)` # 拆分目录名和文件名

分隔

- `os.path.basename(fname)` # 去掉目录路径，返回文件名
- `os.path.dirname(fname)` # 去掉文件名，返回目录路径
- `os.path.join(dirname,basename)` # 组合目录名和文件名
- `os.path.split(fname)` # 拆分目录名和文件名

信息

- `os.path.exists(path or fname)` # 指定路径（文件or目录）
- `os.path.isfile(fname)` 指定路径是否存在且为一个文件
- `os.path.isdir(dirname)` 指定路径是否存在且为一个目录

信息

- `os.path.exists(path or fname)` # 指定路径（文件or目录）
- `os.path.isfile(fname)` 指定路径是否存在且为一个文件
- `os.path.isdir(dirname)` 指定路径是否存在且为一个目录

信息

- `os.path.exists(path or fname)` # 指定路径（文件or目录）
- `os.path.isfile(fname)` 指定路径是否存在且为一个文件
- `os.path.isdir(dirname)` 指定路径是否存在且为一个目录

批量集中文件

- 目标：把一个文件树下全部文件集中到首级文件夹下。
- 思路：
 - 遍历文件树：os.walk
 - 把文件树下每一个文件移动到首级文件夹：for;os.rename
- 实例：sortdata.py
- 拓展：
 - 按拓展名分类文件：if;in
 - 移动到特定文件夹

批量集中文件

- 目标：把一个文件树下全部文件集中到首级文件夹下。
- 思路：
 - 遍历文件树：`os.walk`
 - 把文件树下每一个文件移动到首级文件夹：`for;os.rename`
- 实例：`sortdata.py`
- 拓展：
 - 按拓展名分类文件：`if;in`
 - 移动到特定文件夹

批量集中文件

- 目标：把一个文件树下全部文件集中到首级文件夹下。
- 思路：
 - 遍历文件树：`os.walk`
 - 把文件树下每一个文件移动到首级文件夹：`for;os.rename`
- 实例：`sortdata.py`
- 拓展：
 - 按拓展名分类文件：`if;in`
 - 移动到特定文件夹

批量集中文件

- 目标：把一个文件树下全部文件集中到首级文件夹下。
- 思路：
 - 遍历文件树：os.walk
 - 把文件树下每一个文件移动到首级文件夹：for;os.rename
- 实例：sortdata.py
- 拓展：
 - 按拓展名分类文件：if;in
 - 移动到特定文件夹

批量集中文件

- 目标：把一个文件树下全部文件集中到首级文件夹下。
- 思路：
 - 遍历文件树：os.walk
 - 把文件树下每一个文件移动到首级文件夹：for;os.rename
- 实例：sortdata.py
- 拓展：
 - 按拓展名分类文件：if;in
 - 移动到特定文件夹

批量分类文件

- 目标：把一个文件夹中的文件分成若干500个文件的文件夹。
- 思路：
 - 遍历文件夹下所有文件：for;os.listdir
 - 创建新文件夹：if;os.path.exists;os.mkdir
 - 把文件移动到新文件夹下：os.path.join;os.rename
- 实例：divpdf.py
- 拓展：
 - 按拓展名分类文件：if;in
 - 移动到特定文件夹：if;in

批量分类文件

- 目标：把一个文件夹中的文件分成若干500个文件的文件夹。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `divpdf.py`
- 拓展：
 - 按拓展名分类文件： `if;in`
 - 移动到特定文件夹： `if;in`

批量分类文件

- 目标：把一个文件夹中的文件分成若干500个文件的文件夹。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `divpdf.py`
- 拓展：
 - 按拓展名分类文件： `if;in`
 - 移动到特定文件夹： `if;in`

批量分类文件

- 目标：把一个文件夹中的文件分成若干500个文件的文件夹。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `divpdf.py`
- 拓展：
 - 按拓展名分类文件： `if;in`
 - 移动到特定文件夹： `if;in`

批量分类文件

- 目标：把一个文件夹中的文件分成若干500个文件的文件夹。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `divpdf.py`
- 拓展：
 - 按拓展名分类文件： `if;in`
 - 移动到特定文件夹： `if;in`

批量分类文件

- 目标：把一个文件夹中的文件分成若干500个文件的文件夹。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `divpdf.py`
- 拓展：
 - 按拓展名分类文件： `if;in`
 - 移动到特定文件夹： `if;in`

批量筛选不需要的PDF

- 目标：把英文、已取消、补充公告的财报PDF取出。
- 思路：
 - 创建新文件夹：if;os.path.exists;os.mkdir
 - 遍历文件夹下所有文件：for;os.listdir
 - 判断文件是不是需要筛选出来：if;in
 - 把文件移动到新文件夹下：os.path.join;os.rename
- 实例：selectPassPDF.py

批量筛选不需要的PDF

- 目标：把英文、已取消、补充公告的财报PDF取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 判断文件是不是需要筛选出来： `if;in`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选不需要的PDF

- 目标：把英文、已取消、补充公告的财报PDF取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 判断文件是不是需要筛选出来： `if;in`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选不需要的PDF

- 目标：把英文、已取消、补充公告的财报PDF取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 判断文件是不是需要筛选出来： `if;in`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选不需要的PDF

- 目标：把英文、已取消、补充公告的财报PDF取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 判断文件是不是需要筛选出来： `if;in`
 - 把文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选已处理PDF

- 目标：把已处理PDF（即有对应txt文件的PDF）取出。
- 思路：
 - 创建新文件夹：if;os.path.exists;os.mkdir
 - 遍历TXT文件夹下所有文件：for;os.listdir
 - 判断对应PDF文件是否已经被筛选：str.replace;if,in
 - 把PDF文件移动到新文件夹下：os.path.join;os.rename
- 实例：selectPassPDF.py

批量筛选已处理PDF

- 目标：把已处理PDF（即有对应txt文件的PDF）取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历TXT文件夹下所有文件： `for;os.listdir`
 - 判断对应PDF文件是否已经被筛选： `str.replace;if,in`
 - 把PDF文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选已处理PDF

- 目标：把已处理PDF（即有对应txt文件的PDF）取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历TXT文件夹下所有文件： `for;os.listdir`
 - 判断对应PDF文件是否已经被筛选： `str.replace;if,in`
 - 把PDF文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选已处理PDF

- 目标：把已处理PDF（即有对应txt文件的PDF）取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历TXT文件夹下所有文件： `for;os.listdir`
 - 判断对应PDF文件是否已经被筛选： `str.replace;if;in`
 - 把PDF文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量筛选已处理PDF

- 目标：把已处理PDF（即有对应txt文件的PDF）取出。
- 思路：
 - 创建新文件夹： `if;os.path.exists;os.mkdir`
 - 遍历TXT文件夹下所有文件： `for;os.listdir`
 - 判断对应PDF文件是否已经被筛选： `str.replace;if;in`
 - 把PDF文件移动到新文件夹下： `os.path.join;os.rename`
- 实例： `selectPassPDF.py`

批量提取文件内容

- 目标：批量提取原始文件名中的URL。
- 思路：
 - 打开文件：open
 - 读入文件内容：file.readlines()
 - 从每行提取URL：str.split
 - 写入新文件：open; file.write
- 实例：urlsort.py

批量提取文件内容

- 目标：批量提取原始文件名中的URL。
- 思路：
 - 打开文件： `open`
 - 读入文件内容： `file.readlines()`
 - 从每行提取URL： `str.split`
 - 写入新文件： `open; file.write`
- 实例： `urlsort.py`

批量提取文件内容

- 目标：批量提取原始文件名中的URL。
- 思路：
 - 打开文件： `open`
 - 读入文件内容： `file.readlines()`
 - 从每行提取URL： `str.split`
 - 写入新文件： `open; file.write`
- 实例： `urlsort.py`

批量提取文件内容

- 目标：批量提取原始文件名中的URL。
- 思路：
 - 打开文件： `open`
 - 读入文件内容： `file.readlines()`
 - 从每行提取URL： `str.split`
 - 写入新文件： `open; file.write`
- 实例： `urlsort.py`

批量提取文件内容

- 目标：批量提取原始文件名中的URL。
- 思路：
 - 打开文件： `open`
 - 读入文件内容： `file.readlines()`
 - 从每行提取URL： `str.split`
 - 写入新文件： `open`; `file.write`
- 实例： `urlsort.py`

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件：for;os.listdir
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名：os.path.join;os.rename
- 实例：rename.py

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名： `os.path.join;os.rename`
- 实例： `rename.py`

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名： `os.path.join;os.rename`
- 实例： `rename.py`

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名： `os.path.join;os.rename`
- 实例： `rename.py`

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名： `os.path.join;os.rename`
- 实例： `rename.py`

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名： `os.path.join;os.rename`
- 实例： `rename.py`

批量重命名PDF

- 目标：批量重命名用迅雷下载下来的PDF。
- 思路：
 - 遍历文件夹下所有文件： `for;os.listdir`
 - 生成新旧文件名
 - 读取文件信息
 - 切割每行字符串
 - 提取新旧文件名
 - 把旧文件名重命名成新文件名： `os.path.join;os.rename`
- 实例： `rename.py`

更多

- 从网站上批量下载PDF：爬虫
- 检查TXT文件编码：文件编码
- 统计TXT文件：正则表达式
-

更多

- 从网站上批量下载PDF：爬虫
- 检查TXT文件编码：文件编码
- 统计TXT文件：正则表达式
-

更多

- 从网站上批量下载PDF：爬虫
- 检查TXT文件编码：文件编码
- 统计TXT文件：正则表达式
-

作业

- 实例3-4 基础
- 实例5-6 选做